Journal of Computer and System Sciences 80 (2014) 57–71

Contents lists available at SciVerse ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

Predicting user personality by mining social interactions in Facebook

Alvaro Ortigosa *, Rosa M. Carro, José Ignacio Quiroga

Computer Science Department, Universidad Autónoma de Madrid, Madrid, Spain

ARTICLE INFO

Article history:
Received 25 July 2012
Received in revised form 13 December 2012
Accepted 14 March 2013
Available online 21 March 2013

Keywords:
Data mining in social networks
User modeling
Personality inference

ABSTRACT

Adaptive applications may benefit from having models of users' personality to adapt their behavior accordingly. There is a wide variety of domains in which this can be useful, i.e., assistive technologies, e-learning, e-commerce, health care or recommender systems, among others. The most commonly used procedure to obtain the user personality consists of asking the user to fill in questionnaires. However, on one hand, it would be desirable to obtain the user personality as unobtrusively as possible, yet without compromising the reliability of the model built. On the other hand, our hypothesis is that users with similar personality are expected to show common behavioral patterns when interacting through virtual social networks, and that these patterns can be mined in order to predict the tendency of a user personality. With the goal of inferring personality from the analysis of user interactions within social networks, we have developed TP2010, a Facebook application. It has been used to collect information about the personality traits of more than 20,000 users, along with their interactions within Facebook. Based on all the collected data, automatic classifiers were trained by using different machine-learning techniques, with the purpose of looking for interaction patterns that provide information about the users' personality traits. These classifiers are able to predict user personality starting from parameters related to user interactions, such as the number of friends or the number of wall posts. The results show that the classifiers have a high level of accuracy, making the proposed approach a reliable method for predicting the user personality

1. Motivation

User modeling is essential in the context of adaptive systems. Depending on the goal of each adaptive environment, it can be useful to represent and use some information about the user features, preferences, needs, behaviors, etc. [1]. Regarding the characteristics that can be considered with adaptation purposes, personality is one interesting feature. The personality of an individual can be defined as a set of features that induces a tendency on the behavior of the individual; this tendency is stable through time and situations [2]. Knowing the personality of a given person provides hints about how he would probably react when facing different situations.

Identifying a user's personality can contribute to know, for example, his potential needs in different contexts. Therefore, adaptive applications may benefit from having models of users' personality to adapt their behavior accordingly. There is a wide variety of domains in which this can be useful, i.e., assistive technologies, e-learning, e-commerce, health care or recommender systems, among others. For example, in the context of e-commerce, the type of products to be offered to a user may vary depending on his personality with respect to *Impulsive Sensation Seeking*. Another application area would

* Corresponding author.
E-mail addresses: alvaro.ortigosa@uam.es (A. Ortigosa), rosa.carro@uam.es (R.M. Carro), jose.quiroga@estudiante.uam.es (J.I. Quiroga).

0022-0000/$ – see front matter
http://dx.doi.org/10.1016/j.jcss.2013.03.008

be education, since personality influences the way in which students learn and use their knowledge [3], and affects the way in which students accomplish their tasks [4]. In adaptive educational systems it would be useful to know the student personality to propose him the most suitable tasks in each context accordingly (for example, if he has only 10 minutes available and connects to the system, it could be inappropriate to propose him to develop a complex task, even if it is not so time-consuming, if he has a high level of *Neuroticism-Anxiety*).

Therefore, eliciting user personality can contribute significantly to user modeling for adaptive systems. The most commonly used procedure to obtain this information consists of asking the user to fill in questionnaires. However, users can find this task too time-consuming, since most of the personality questionnaires include many questions to answer in order to obtain an accurate user profile [5,6].

On one hand, we think that user personality should be obtained as unobtrusively as possible, yet without compromising the reliability of the model built. Some research has already been done related to automatic or semi-automatic user modeling acquisition. Several works have tried to infer some personality characteristics, as it will be discussed in the state of the art section.

On the other hand, we think that personality can be inferred by analyzing how users interact in online social networks (OSNs). In this context, an initial consideration would be whether virtual interactions reflect user personality in "real" or offline life. For example, [7] and [8] verify that Facebook,[1] currently the most popular OSN, is, apparently, a good approximation to the user offline life. If a given user has occasional interactions with another user through the Web, it will not be the case that they have many more offline interactions. In addition, it has been shown that people use virtual social networks to support already existing relationships rather than looking for new ones (for Facebook users, around 77% of their social relationships in the real world are replicated in the virtual environment).

In the same line, the user interactions on social networks are a good measure of user behavior in real life [9]. They verify this hypothesis analyzing user comments on pictures and publications of Facebook wall. For example, they show that even if a user has a lot of relations ("friends" in Facebook terminology) he only interacts regularly with a small portion of them. That is because there are, like in real life, constraints that make interacting with all friends impossible (e.g., available time).

Furthermore, [10] shows that, contrary to the idea that people built idealized virtual-identity, Facebook profiles reflect actual personality. The same idea is supported by [11], where it was found out that personality impression based on Facebook profiles generally show strong patterns of convergence with real personality.

However, it is important to bear in mind that this work does not presume that users would act exactly in the same way both in real and in virtual life. What is known is that personality, by definition, is stable through time and situations. That means that if a user has tendency to be, for example, sociable, that tendency would show up both on real and virtual life, maybe in different ways. What this work assumes is that users with similar personality features will show common behavior patterns when interacting through virtual social networks, that those patterns can be mined, and that they can be used, afterwards, to identify the tendency of a user personality. In this direction, we think that it may be possible to infer personality from the analysis of user interactions within online social networks, such as the number of user's friends or the number of wall posts.

There exist a growing number of online social networks available through the Web. From these applications, Facebook is by large the more popular around the world. On June 2012, it reached 955 million of active users (that is, without considering the profiles created but not used) [12]. Fig. 1a shows the growing rate: in the last three months it increased 54 million users (almost 6%). More than a half of Facebook users are daily active users (552 millions), as shown in Fig. 1b.

Fig. 2 also shows that more than 543 million users access and update their profiles through mobile devices. Besides, according to [13], Facebook users have an average of 130 "friends" (the term used within Facebook to denote people related to the user) and are connected, on the average, to 60 items, counting groups, pages and events. The Facebook application was translated to more than 70 languages; it has more than one million of developers and more than 500,000 available applications.

Among all the social networks available through the Internet nowadays, we have chosen Facebook not only because of the huge amount of users connected to it, but also because it offers the possibility of building applications to be incorporated and used by all these users. In this context, we have developed TP2010 [14], a Facebook application intended to obtain information about the user personality through a personality test, as well as to collect all the data available from the user interactions within the social network. The application was delivered on December 2009, and it implements the ZKPQ-50-cc personality test [5].

TP2010 goal is to discover relationships between the users' results on the personality test and those attributes describing their interactions within Facebook. Our work attempts to find rules to predict user personality (that is, the tendency of the user on each personality dimension being considered), starting from data about his interactions in Facebook, without asking him to fulfill specific personality tests. With this aim, we have applied data mining techniques in order to build classifiers of user personality. These classifiers were trained based on the analysis of data from more than 20,000 users of the TP2010 application. Most of the related works found in the literature that try to find relationships between user personality and user interactions in social networks focus on analyzing how single features correlates, on the average, with personality traits. We intend to go beyond and analyze the collected data by looking for patterns of user interactions that correspond to

[1] www.facebook.com.